Chapter 2 - Axioms of Probability

2.1 - Introduction

• Introduce concept of probability of an event and then show how probabilities can be computed.

* Preliminary: concept of sample space & events of an event

2.2 - Sample space & events

- The set of all possible outcomes of an experiment is known as sample space. [S]
- Any subset [E] of the sample space is known as event
  - eg $(E \cup F)$ = all outcomes E & F
  - eg $(E \cap F)$ = all outcomes both E, F.
  - eg $(EF = \emptyset)$ = mutually exclusive

[event consisting of no outcomes]

→ union: $\bigcup_{n=1}^{\infty} E_n$

→ intersection: $\bigcap_{n=1}^{\infty} E_n$

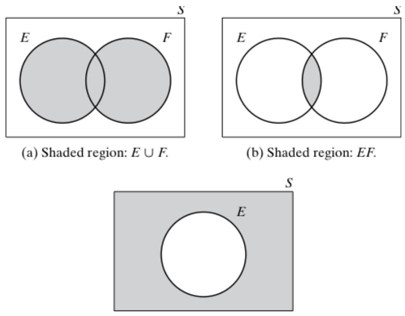
→ E^c (complement of E)

↳ all outcomes in sample space S that are <u>not</u> in E.

<u>Note</u>: That because the experiment must result in some outcome, it follows that $S^c = \phi$

→ $E \subset F$ (subset)

All outcomes of E are in also F.



(a) Shaded region: $E \cup F$. (b) Shaded region: EF.

(c) Shaded region: E^c.

FIGURE 2.1: Venn Diagrams

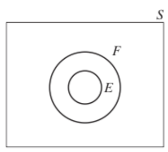
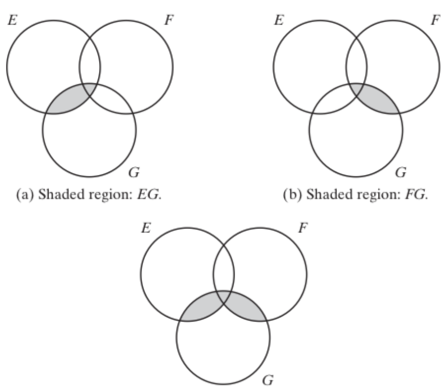


FIGURE 2.2: $E \subset F$

The operations of forming unions, intersections, and complements of events obey certain rules similar to the rules of algebra. We list a few of these rules:

Commutative laws	$E \cup F = F \cup E$	$EF = FE$
Associative laws	$(E \cup F) \cup G = E \cup (F \cup G)$	$(EF)G = E(FG)$
Distributive laws	$(E \cup F)G = EG \cup FG$	$EF \cup G = (E \cup G)(F \cup G)$

These relations are verified by showing that any outcome that is contained in the event on the left side of the equality sign is also contained in the event on the right side, and vice versa. One way of showing this is by means of Venn diagrams. For instance, the distributive law may be verified by the sequence of diagrams in Figure 2.3.



(a) Shaded region: EG. (b) Shaded region: FG.

(c) Shaded region: $(E \cup F)G$.

FIGURE 2.3: $(E \cup F)G = EG \cup FG$

The following useful relationships between the three basic operations of forming unions, intersections, and complements are known as *DeMorgan's laws*:

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

To prove DeMorgan's laws, suppose first that x is an outcome of $\left(\bigcup\limits_{i=1}^{n} E_i\right)^c$. Then x is not contained in $\bigcup\limits_{i=1}^{n} E_i$, which means that x is not contained in any of the events $E_i, i = 1, 2, \ldots, n$, implying that x is contained in E_i^c for all $i = 1, 2, \ldots, n$ and thus is contained in $\bigcap\limits_{i=1}^{n} E_i^c$. To go the other way, suppose that x is an outcome of $\bigcap\limits_{i=1}^{n} E_i^c$. Then x is contained in E_i^c for all $i = 1, 2, \ldots, n$, which means that x is not contained in E_i for any $i = 1, 2, \ldots, n$, implying that x is not contained in $\bigcup\limits_{i}^{n} E_i$, in turn implying that x is contained in $\left(\bigcup\limits_{1}^{n} E_i\right)^c$. This proves the first of DeMorgan's laws.

To prove the second of DeMorgan's laws, we use the first law to obtain

$$\left(\bigcup_{i=1}^{n} E_i^c\right)^c = \bigcap_{i=1}^{n} (E_i^c)^c$$

which, since $(E^c)^c = E$, is equivalent to

$$\left(\bigcup_{1}^{n} E_i^c\right)^c = \bigcap_{1}^{n} E_i$$

Taking complements of both sides of the preceding equation yields the result we seek, namely,

$$\bigcup_{1}^{n} E_i^c = \left(\bigcap_{1}^{n} E_i\right)^c$$

2.3- Axioms of Probability

One way of defining the probability of an event is in terms of relative frequency.

→ We suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions

- For each event E of the sample space S, we define $n(E)$ to be the number of times in the first n-repetitions of the experiment that the event E occurs.

(=)

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \Bigg\} \text{Probability of event } E$$

↳ Probability is defined as the (limiting) proportion of time that E occurs.
[limiting frequency of E]

Note: Although this definition is intuitive

Problem: How do we know that $n(E)/n$ will converge to some constant limiting factor that will be the same for each possible sequence of repetitions of the experiment?

Eg: suppose flipping a coin experiment.
→ How do we know that the proportion of heads obtained in the first n flips will converge to some value as $n >>$

Proponents of the relative frequency definition of probability answers these objections by stating that the convergence of $n(E)/n$ to a constant limiting value is an assumption / axiom of the system

Axiom 1

$$0 \le P(E) \le 1$$

Axiom 2

$$P(S) = 1$$

Axiom 3

For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i E_j = \emptyset$ when $i \ne j$),

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

We refer to $P(E)$ as the probability of the event E.

Thus, Axiom 1 states that the probability that the outcome of the experiment is an outcome in E is some number between 0 and 1. Axiom 2 states that, with probability 1, the outcome will be a point in the sample space S. Axiom 3 states that, for any sequence of mutually exclusive events, the probability of at least one of these events occurring is just the sum of their respective probabilities.

If we consider a sequence of events $E_1, E_2, \ldots$, where $E_1 = S$ and $E_i = \emptyset$ for $i > 1$, then, because the events are mutually exclusive and because $S = \bigcup_{i=1}^{\infty} E_i$, we have, from Axiom 3,

$$P(S) = \sum_{i=1}^{\infty} P(E_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

implying that

$$P(\emptyset) = 0$$

That is, the null event has probability 0 of occurring.

Note that it follows that, for any finite sequence of mutually exclusive events E_1, $E_2, \ldots, E_n$,

$$P\left(\bigcup_{1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \tag{3.1}$$

This equation follows from Axiom 3 by defining E_i as the null event for all values of i greater than n. Axiom 3 is equivalent to Equation (3.1) when the sample space is finite. (Why?) However, the added generality of Axiom 3 is necessary when the sample space consists of an infinite number of points.

EXAMPLE 3b

If a die is rolled and we suppose that all six sides are equally likely to appear, then we would have $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$. From Axiom 3, it would thus follow that the probability of rolling an even number would equal

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2} \qquad \blacksquare$$

The assumption of the existence of a set function P, defined on the events of a sample space S and satisfying Axioms 1, 2, and 3, constitutes the modern mathematical approach to probability theory. Hopefully, the reader will agree that the axioms are natural and in accordance with our intuitive concept of probability as related to chance and randomness. Furthermore, using these axioms we shall be able to prove that if an experiment is repeated over and over again, then, with probability 1, the proportion of time during which any specific event E occurs will equal $P(E)$. This result, known as the strong law of large numbers, is presented in Chapter 8. In addition, we present another possible interpretation of probability—as being a measure of belief—in Section 2.7.

Technical Remark. We have supposed that $P(E)$ is defined for all the events E of the sample space. Actually, when the sample space is an uncountably infinite set, $P(E)$ is defined only for a class of events called measurable. However, this restriction need not concern us, as all events of any practical interest are measurable.

2.4 - Some Simple Propositions

Proof of Propositions regarding probabilities.

Note:

Since E & E^c are mutually exclusive and since $E \cup E^c = S$ we have Axioms 2 & 3

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

Proposition 4.1.

$$P(E^c) = 1 - P(E)$$

In words, Proposition 4.1 states that the probability that an event does not occur is 1 minus the probability that it does occur. For instance, if the probability of obtaining a head on the toss of a coin is $\frac{3}{8}$, then the probability of obtaining a tail must be $\frac{5}{8}$.

Our second proposition states that if the event E is contained in the event F, then the probability of E is no greater than the probability of F.

Proposition 4.2. If $E \subset F$, then $P(E) \leq P(F)$.

Proof. Since $E \subset F$, it follows that we can express F as

$$F = E \cup E^c F$$

Hence, because E and $E^c F$ are mutually exclusive, we obtain, from Axiom 3,

$$P(F) = P(E) + P(E^c F)$$

which proves the result, since $P(E^c F) \geq 0$. □

Proposition 4.2 tells us, for instance, that the probability of rolling a 1 with a die is less than or equal to the probability of rolling an odd value with the die.

The next proposition gives the relationship between the probability of the union of two events, expressed in terms of the individual probabilities, and the probability of the intersection of the events.

Proposition 4.3.

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Proof. To derive a formula for $P(E \cup F)$, we first note that $E \cup F$ can be written as the union of the two disjoint events E and E^cF. Thus, from Axiom 3, we obtain

$$\begin{aligned} P(E \cup F) &= P(E \cup E^cF) \\ &= P(E) + P(E^cF) \end{aligned}$$

Furthermore, since $F = EF \cup E^cF$, we again obtain from Axiom 3

$$P(F) = P(EF) + P(E^cF)$$

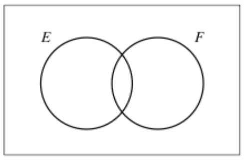


FIGURE 2.4: Venn Diagram

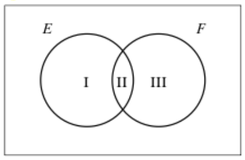


FIGURE 2.5: Venn Diagram in Sections

or, equivalently,

$$P(E^cF) = P(F) - P(EF)$$

thereby completing the proof. □

Proposition 4.3 could also have been proved by making use of the Venn diagram in Figure 2.4.

Let us divide $E \cup F$ into three mutually exclusive sections, as shown in Figure 2.5. In words, section I represents all the points in E that are not in F (that is, EF^c), section II represents all points both in E and in F (that is, EF), and section III represents all points in F that are not in E (that is, E^cF).

From Figure 2.5, we see that

$$\begin{aligned} E \cup F &= \text{I} \cup \text{II} \cup \text{III} \\ E &= \text{I} \cup \text{II} \\ F &= \text{II} \cup \text{III} \end{aligned}$$

As I, II, and III are mutually exclusive, it follows from Axiom 3 that

$$\begin{aligned} P(E \cup F) &= P(\text{I}) + P(\text{II}) + P(\text{III}) \\ P(E) &= P(\text{I}) + P(\text{II}) \\ P(F) &= P(\text{II}) + P(\text{III}) \end{aligned}$$

which shows that

$$P(E \cup F) = P(E) + P(F) - P(\text{II})$$

and Proposition 4.3 is proved, since $\text{II} = EF$.

Inclusion - exclusion identity (Proposition 4.4) can be proved by mathematical induction (- IMPORTANT.)

Proposition 4.4.

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots$$
$$+ (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$$
$$+ \cdots + (-1)^{n+1} P(E_1E_2\cdots E_n)$$

The summation $\sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$ is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \ldots, n\}$.

In words, Proposition 4.4 states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time, minus the sum of the probabilities of these events taken two at a time, plus the sum of the probabilities of these events taken three at a time, and so on.

Remarks. 1. For a noninductive argument for Proposition 4.4, note first that if an outcome of the sample space is not a member of any of the sets E_i, then its probability does not contribute anything to either side of the equality. Now, suppose that an outcome is in exactly m of the events E_i, where $m > 0$. Then, since it is in $\bigcup_i E_i$, its

probability is counted once in $P\left(\bigcup_i E_i\right)$; also, as this outcome is contained in $\binom{m}{k}$ subsets of the type $E_{i_1}E_{i_2}\cdots E_{i_k}$, its probability is counted

$$\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \cdots \pm \binom{m}{m}$$

times on the right of the equality sign in Proposition 4.4. Thus, for $m > 0$, we must show that

$$1 = \binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \cdots \pm \binom{m}{m}$$

However, since $1 = \binom{m}{0}$, the preceding equation is equivalent to

$$\sum_{i=0}^{m} \binom{m}{i} (-1)^i = 0$$

and the latter equation follows from the binomial theorem, since

$$0 = (-1 + 1)^m = \sum_{i=0}^{m} \binom{m}{i} (-1)^i (1)^{m-i}$$

2. The following is a succinct way of writing the inclusion–exclusion identity:

$$P(\cup_{i=1}^{n} E_i) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<\cdots<i_r} P(E_{i_1}\cdots E_{i_r})$$

3. In the inclusion–exclusion identity, going out one term results in an upper bound on the probability of the union, going out two terms results in a lower bound on the probability, going out three terms results in an upper bound on the probability, going out four terms results in a lower bound, and so on. That is, for events $E_1, \ldots, E_n$, we have

$$P(\cup_{i=1}^n E_i) \le \sum_{i=1}^n P(E_i) \tag{4.1}$$

$$P(\cup_{i=1}^n E_i) \ge \sum_{i=1}^n P(E_i) - \sum_{j<i} P(E_i E_j) \tag{4.2}$$

$$P(\cup_{i=1}^n E_i) \le \sum_{i=1}^n P(E_i) - \sum_{j<i} P(E_i E_j) + \sum_{k<j<i} P(E_i E_j E_k) \tag{4.3}$$

and so on. To prove the validity of these bounds, note the identity

$$\cup_{i=1}^n E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \cdots \cup E_1^c \cdots E_{n-1}^c E_n$$

That is, at least one of the events E_i occurs if E_1 occurs, or if E_1 does not occur but E_2 does, or if E_1 and E_2 do not occur but E_3 does, and so on. Because the right-hand side is the union of disjoint events, we obtain

$$\begin{aligned} P(\cup_{i=1}^n E_i) &= P(E_1) + P(E_1^c E_2) + P(E_1^c E_2^c E_3) + \ldots + P(E_1^c \cdots E_{n-1}^c E_n) \\ &= P(E_1) + \sum_{i=2}^n P(E_1^c \cdots E_{i-1}^c E_i) \end{aligned} \tag{4.4}$$

Now, let $B_i = E_1^c \cdots E_{i-1}^c = (\cup_{j<i} E_j)^c$ be the event that none of the first $i-1$ events occur. Applying the identity

$$P(E_i) = P(B_i E_i) + P(B_i^c E_i)$$

shows that

$$P(E_i) = P(E_1^c \cdots E_{i-1}^c E_i) + P(E_i \cup_{j<i} E_j)$$

or, equivalently,

$$P(E_1^c \cdots E_{i-1}^c E_i) = P(E_i) - P(\cup_{j<i} E_i E_j)$$

Substituting this equation into (4.4) yields

$$P(\cup_{i=1}^n E_i) = \sum_i P(E_i) - \sum_i P(\cup_{j<i} E_i E_j) \tag{4.5}$$

Because probabilities are always nonnegative, Inequality (4.1) follows directly from Equation (4.5). Now, fixing i and applying Inequality (4.1) to $P(\cup_{j<i} E_i E_j)$ yields

$$P(\cup_{j<i} E_i E_j) \le \sum_{j<i} P(E_i E_j)$$

which, by Equation (4.5), gives Inequality (4.2). Similarly, fixing i and applying Inequality (4.2) to $P(\cup_{j<i} E_i E_j)$ yields

$$\begin{aligned} P(\cup_{j<i} E_i E_j) &\ge \sum_{j<i} P(E_i E_j) - \sum_{k<j<i} P(E_i E_j E_i E_k) \\ &= \sum_{j<i} P(E_i E_j) - \sum_{k<j<i} P(E_i E_j E_k) \end{aligned}$$

which, by Equation (4.5), gives Inequality (4.3). The next inclusion–exclusion inequality is now obtained by fixing i and applying Inequality (4.3) to $P(\cup_{j<i} E_i E_j)$, and so on.

2.5 - Sample spaces having equally likely outcomes

- We can sometimes assume that all outcomes in the sample space are equally likely to occur.

$\therefore$ from Axiom 1 & 2

$$P(\{1\}) = P(\{2\}) = \dots = P(\{N\})$$

$$\therefore P(\{i\}) = \frac{1}{N} \quad \text{where } i = 1, 2, \dots, N$$

$\therefore$ Following Axiom 3, any event E

$$\boxed{P(E) = \frac{\#\text{ of outcomes } E}{\#\text{ of outcomes } S}}$$

(* Example found pg 49-57)

*2.6 : Probability as a continuous set function * Very important proofs

*2.6 PROBABILITY AS A CONTINUOUS SET FUNCTION

A sequence of events $\{E_n, n \geq 1\}$ is said to be an increasing sequence if

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

whereas it is said to be a decreasing sequence if

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

If $\{E_n, n \geq 1\}$ is an increasing sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$, by

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

Similarly, if $\{E_n, n \geq 1\}$ is a decreasing sequence of events, we define $\lim_n E_n$ by

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

We now prove the following Proposition 1:

Proposition 6.1.
If $\{E_n, n \geq 1\}$ is either an increasing or a decreasing sequence of events, then

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

Proof. Suppose, first, that $\{E_n, n \geq 1\}$ is an increasing sequence, and define the events $F_n, n \geq 1$, by

$$F_1 = E_1$$
$$F_n = E_n\left(\bigcup_1^{n-1} E_i\right)^c = E_n E_{n-1}^c \quad n > 1$$

where we have used the fact that $\bigcup_1^{n-1} E_i = E_{n-1}$, since the events are increasing. In words, F_n consists of those outcomes in E_n which are not in any of the earlier $E_i, i < n$. It is easy to verify that the F_n are mutually exclusive events such that

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \quad \text{and} \quad \bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \quad \text{for all } n \geq 1$$

Thus,

$$\begin{aligned}
P\left(\bigcup_1^{\infty} E_i\right) &= P\left(\bigcup_1^{\infty} F_i\right) \\
&= \sum_1^{\infty} P(F_i) \quad \text{(by Axiom 3)} \\
&= \lim_{n\to\infty} \sum_1^{n} P(F_i) \\
&= \lim_{n\to\infty} P\left(\bigcup_1^{n} F_i\right) \\
&= \lim_{n\to\infty} P\left(\bigcup_1^{n} E_i\right) \\
&= \lim_{n\to\infty} P(E_n)
\end{aligned}$$

which proves the result when $\{E_n, n \geq 1\}$ is increasing.

If $\{E_n, n \geq 1\}$ is a decreasing sequence, then $\{E_n^c, n \geq 1\}$ is an increasing sequence hence, from the preceding equations,

$$P\left(\bigcup_1^{\infty} E_i^c\right) = \lim_{n\to\infty} P(E_n^c)$$

$$P\left(\bigcup_1^\infty E_i^c\right) = \lim_{n\to\infty} P(E_n^c)$$

However, because $\bigcup_1^\infty E_i^c = \left(\bigcap_1^\infty E_i\right)^c$, it follows that

$$P\left(\left(\bigcap_1^\infty E_i\right)^c\right) = \lim_{n\to\infty} P(E_n^c)$$

or, equivalently,

$$1 - P\left(\bigcap_1^\infty E_i\right) = \lim_{n\to\infty}[1 - P(E_n)] = 1 - \lim_{n\to\infty} P(E_n)$$

or

$$P\left(\bigcap_1^\infty E_i\right) = \lim_{n\to\infty} P(E_n)$$

which proves the result. □

* Example 6a
Probability & a paradox

→ Lets go through it

- Suppose that we have an infinitely large urn and an infinite collection of balls labelled ball number 1, 2, 3 .... n.

• Consider an experiment performed as follows.

- At 1 minute to 12 PM, ball numbered 1 through 10 are placed in the urn and ball number 10 is withdrawn.
(Assume that withdrawal takes no time).

- At 1/2 minute to 12 pm., balls numbered 11 through 20 are placed in the urn and ball number 20 is withdrawn.

- At 1/4 minute to 12 pm, balls, numbered 21 to 30 are placed in the urn and ball number 30 is withdrawn.

- At 1/8 minute to 12 pm, and so on ...

Question: How many balls are in the urn at 12 pm?

→ The answer to this question is clearly that there is an infinite number of balls in the urn at 12 pm, since any ball whose number is not in the form $10n$, $n \geq 1$, will have been placed in the urn & will not have been withdrawn before 12 pm.

⇒ Problem solved when experiment is performed as described.

Changing the problem

- Suppose that at 1 minute to 12pm balls number 1-10 are placed in the urn and ball numbered 1 is withdrawn
- At 1/2 minute before 12pm ball numbered 11-20 are placed in the urn and ball number 2 is is withdrawn
- At 1/4 minute before 12pm balls numbered 21-30 are placed in the urn and ball # 3 is withdrawn

⇒ Problem goes on in similar fashion

Question: For this new experiment how many balls are in the urn at 12pm?

* Surprisingly the answer now is that the urn is empty at 12pm.

→ Consider any ball, say ball number n.
At some point to prior to 12pm [in particular at $(1/2)^{n-1}$ minutes to 12pm], this ball would have been withdrawn from the urn.
Hence, for each n, ball number n is not in the urn at 12pm.

Therefore the urn must be empty at that time.

* We see then, that the manner in which the balls are withdrawn makes a difference.

Change in problem:
→ Consider now that at 1 minute to 12 pm, balls numbered 1 to 10 are placed in the urn and a ball is randomly selected and withdrawn

Question: How many balls are in the urn at 12PM?

⇒ We shall show that, with probability 1, the urn is empty at 12 pm.
↳ Let's first consider ball number 1.
→ Define E_n, to be the event that ball number 1 is still in the urn after the first n withdrawals have been made

$$\mathbb{P}(E_n) = \frac{9 \cdot 18 \cdot 27 \cdot \cdots (9n)}{10 \cdot 19 \cdot 28 \cdots (9n+1)}$$

⇒ Understand this equation
Note that if ball number 1 ~~is~~ still to be in the urn after

the first $\underline{n}$ withdrawals, the first ball withdrawn can be any one of 9, the second any one of 18 (there are 19 balls in the urn at the time of second withdrawal, one of which must be ball number 1.

[same for denominator]

$\Rightarrow$ Now, the event that the ball number 1 is in the urn at 12pm, is just the event

$\bigcap_{n=1}^{\infty} E_n$ (decreasing function)

Because the events $E_n, n \geq 1$, are decreasing events, it follows from Proposition 6.1 that

$$\begin{aligned}
P\{&\text{ball number 1 is in the urn at 12 P.M.}\} \\
&= P\left(\bigcap_{n=1}^{\infty} E_n\right) \\
&= \lim_{n\to\infty} P(E_n) \\
&= \prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right)
\end{aligned}$$

We now show that

$$\prod_{n=1}^{\infty}\frac{9n}{9n+1} = 0$$

Since

$$\prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right) = \left[\prod_{n=1}^{\infty}\left(\frac{9n+1}{9n}\right)\right]^{-1}$$

this is equivalent to showing that

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) = \infty$$

Now, for all $m \geq 1$,

$$\begin{aligned}
\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) &\geq \prod_{n=1}^{m}\left(1+\frac{1}{9n}\right) \\
&= \left(1+\frac{1}{9}\right)\left(1+\frac{1}{18}\right)\left(1+\frac{1}{27}\right)\cdots\left(1+\frac{1}{9m}\right) \\
&> \frac{1}{9}+\frac{1}{18}+\frac{1}{27}+\cdots+\frac{1}{9m} \\
&= \frac{1}{9}\sum_{i=1}^{m}\frac{1}{i}
\end{aligned}$$

Hence, letting $m\to\infty$ and using the fact that $\sum_{i=1}^{\infty} 1/i = \infty$ yields

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) = \infty$$

Thus, letting F_i denote the event that ball number i is in the urn at 12 P.M., we have shown that $P(F_1) = 0$. Similarly, we can show that $P(F_i) = 0$ for all i.

For instance:

(For instance, the same reasoning shows that $P(F_i) = \prod_{n=2}^{\infty}[9n/(9n+1)]$ for $i = 11, 12, \ldots, 20$.) Therefore, the probability that the urn is not empty at 12 P.M., $P\left(\bigcup_1^{\infty} F_i\right)$, satisfies

$$P\left(\bigcup_1^{\infty} F_i\right) \leq \sum_1^{\infty} P(F_i) = 0$$

by Boole's inequality. (See Self-Test Exercise 14.)

Thus, with probability 1, the urn will be empty at 12 P.M. ■

2.7 - Probability as a measure of belief

→ It is logical to suppose that "measure of the degree of one's belief", should satisfy all of the axioms of probability.

Hence we can interpret probability as a measure of belief or as

a long-run frequency of occurance
its mathematical properties
remain the same.

SUMMARY

Let S denote the set of all possible outcomes of an experiment. S is called the *sample space* of the experiment. An event is a subset of S. If $A_i, i = 1, \ldots, n$, are events, then $\bigcup_{i=1}^{n} A_i$, called the *union* of these events, consists of all outcomes that are in at least one of the events $A_i, i = 1, \ldots, n$. Similarly, $\bigcap_{i=1}^{n} A_i$, sometimes written as $A_1 \cdots A_n$, is called the *intersection* of the events A_i and consists of all outcomes that are in all of the events $A_i, i = 1, \ldots, n$.

For any event A, we define A^c to consist of all outcomes in the sample space that are not in A. We call A^c the *complement* of the event A. The event S^c, which is empty of outcomes, is designated by $\emptyset$ and is called the *null* set. If $AB = \emptyset$, then we say that A and B are *mutually exclusive*.

For each event A of the sample space S, we suppose that a number $P(A)$, called the probability of A, is defined and is such that

(i) $0 \leq P(A) \leq 1$
(ii) $P(S) = 1$
(iii) For mutually exclusive events $A_i, i \geq 1$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$P(A)$ represents the probability that the outcome of the experiment is in A.

It can be shown that

$$P(A^c) = 1 - P(A)$$

A useful result is that

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

which can be generalized to give

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}\sum P(A_iA_j) + \sum_{i<j<k}\sum\sum P(A_iA_jA_k) + \cdots + (-1)^{n+1}P(A_1 \cdots A_n)$$

If S is finite and each one point set is assumed to have equal probability, then

$$P(A) = \frac{|A|}{|S|}$$

where $|E|$ denotes the number of outcomes in the event E.

$P(A)$ can be interpreted either as a long-run relative frequency or as a measure of one's degree of belief.